確率最適化における過去集積値と未来閾値について (不確実なモデルによる動的計画理論の課題とその展望)

AUTHOR(S):

植野, 貴之; 岩本, 誠一

CITATION:

植野, 貴之 ...[et al]. 確率最適化における過去集積値と未来閾値について (不確実なモデルによる動的計画理論の課題とその展望). 数理解析研究所講究録 2001, 1207: 79-100

ISSUE DATE:

2001-05

URL:

http://hdl.handle.net/2433/41045

数理解析研究所講究録 1207 巻 2001 年 79-100

確率最適化における過去集積値と未来閾値について

九大大学院経済学研究科　植野貴之 (Takayuki Ueno)
Graduate School of Economics, Kyushu University
九大大学院経済学研究院　岩本誠一 (Seiichi Iwamoto)
Faculty of Economics, Kyushu University

1　はじめに

本論文では、不確実性の下で多段階にわたる加法型（総合）評価値が所定の基準値以上になる（閾値）確率を最適化する問題を考える。2種類の－（1）過去集積値に基づく（2）未来閾値に基づく－拡大マルコフ政策を導入して、それぞれのクラスでの最適政策を求める。さらに、両政策の間に双対性および一致性が成り立つことを示す。期待値最適化問題では、最適政策はマルコフ政策クラスの中に存在することがわかっているが、閾値確率最適化問題では一般にマルコフ政策は十分でなく、広く一般政策クラスに最適政策が存在することを示す。

2　閾値確率制御問題

以下、本論文で用いる記号と用語を述べておこう。

(1)　$N \geq 2$ は**段の総数** (total number of stage) を表す正整数

(2)　$X = \{s_1, s_2, \ldots, s_p\}$ は有限**状態空間** (state space)

(3)　$U = \{a_1, a_2, \ldots, a_k\}$ は有限**決定空間** (action space)

(4)　$r_n : X \times U \to R^1$ は**第 n 利得関数** (n−th reward function)　$(0 \leq n \leq N-1)$

; $r_n(x,u)$ は第 n 期に状態 x で決定 u をとったとき、システムから得られる利得（リターン）を表す。

$r_N : X \to R^1$ は**終端利得関数** (terminal reward function)

; $r_N(x)$ は最終 N 期に状態 x になったとき、システムから得られる終端利得を表す。

(5)　$p = \{p(y|x,u)\}$ は**マルコフ推移法則** (Markov transition law)

: $p(y|x,u) \geq 0 \quad \forall (x,u,y) \in X \times U \times X$

$\sum_{y \in X} p(y|x,u) = 1 \quad \forall (x,u) \in X \times U$

; $p(y|x,u)$ は現在状態 $\tilde{X}$ が x で現在の決定 $\tilde{U}$ が u になったとき、次の状態 $\tilde{Y}$ が y になる条件付き確率を表す：$P(\tilde{Y} = y | \tilde{X} = x, \tilde{U} = u) = p(y|x,u)$. ただし $\tilde{}$ は確率変数を表す。この確率的推移を $\tilde{Y} \sim p(\cdot|x,u)$ で表現する。

(6) $\pi=\{\pi_0, \pi_1, \ldots, \pi_{N-1}\}$ は**マルコフ政策** (Markov policy)
: $\pi_0 : X \to U, \quad \pi_1 : X \to U, \quad \ldots \quad , \quad \pi_{N-1} : X \to U$
第 n 段までに状態列 $x_0, x_1, \ldots, x_n$ を経てきたとき、意志決定者は途中の状態列 $x_0, x_1, \ldots, x_{n-1}$ に無関係に決定 $\pi_n(x_n) \in U$ を取ることを表している。マルコフ政策の全体を Π で表す。

(6)′ $\sigma=\{\sigma_0, \sigma_1, \ldots, \sigma_{N-1}\}$ は**一般政策** (general policy)
: $\sigma_0 : X \to U, \quad \sigma_1 : X \times X \to U, \quad \ldots \quad , \quad \sigma_{N-1} : X \times \cdots \times X \to U$
第 n 段までに状態列 $x_0, x_1, \ldots, x_n$ を観察したとき、意志決定者は決定 $\sigma_n(x_0, x_1, \ldots, x_n) \in U$ を取ることを表している。一般政策の全体を $\Pi(g)$ で表す。

(6)″ $\mu=\{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$ は**原始政策** (primitive policy)
: $\mu_0 : X \to U, \quad \mu_1 : X \times U \times X \to U, \quad \ldots$
$\quad , \quad \mu_{N-1} : X \times U \times X \times U \times \cdots \times U \times X \to U$
第 n 段までに状態と決定の交互列 $x_0, u_0, x_1, u_1, \ldots, u_{n-1}, x_n$ を経てきたとき、意志決定者は途中の決定列 $u_0, u_1, \ldots, u_{n-1}$ にも依存して決定 $\sigma_n(x_0, u_0, x_1, u_1, \ldots, u_{n-1}, x_n) \in U$ を取ることを表している。原始政策の全体を $\Pi(p)$ で表す。

いま、ある N 段システムが、初期（第0段）状態 $x_0 \in X$ から出発して制御マルコフ推移法則 $p=\{p(y|x,u)\}$ に従って推移し、決定と状態の交互列

$$u_0 \in U,\ x_1 \in X,\ u_1 \in U,\ x_2 \in X,\ \ldots, u_{N-1} \in U$$

を経て、最終的にある確率で $x_N \in X$ になり、そこで終了するとする。このとき、第0段では状態 x_0 と決定 u_0 に依存した利得（リターン） $r_0(x_0, u_0)$ が得られ、第1段では x_1 と u_1 に関係した利得 $r_1(x_1, u_1)$ が得られ、以下、第 n 段では利得 $r_n(x_n, u_n)$ が得られ、最終の第 N 段終了時点には、さらに終端状態 x_N に依存した終端利得 $r_N(x_N)$ が得られるとする。すなわち、意志決定者はシステム全体を通じては各段で得られた利得の総和

$$r_0(x_0, u_0) + r_1(x_1, u_1) + \cdots + r_{N-1}(x_{N-1}, u_{N-1}) + r_N(x_N)$$

を得るものとする。不確実な状況の下でこの総和があらかじめ定められた（所定の）値 c 以上になる確率を最大にするように行動するには、意志決定者が各段で、それまでの状態に応じてどのように決定を取っていけばよいかが問題である。一般に、第 n 期の状態と決定をそれぞれ確率変数 X_n, U_n で表わすと、得られる総利得は確率変数

$$r_0(X_0, U_0) + r_1(X_1, U_1) + \cdots + r_{N-1}(X_{N-1}, U_{N-1}) + r_N(X_N)$$

で表わされる。総利得を表わす確率変数を以後簡単に

$$r_0 + r_1 + \cdots + r_{N-1} + r_N$$

で表す。したがって、問題を数学的に記述すると、次の閾値確率最大化問題になる：

$$\mathrm{P}_0(x_0)\qquad \begin{array}{ll} \text{Maximize} & P^{\sigma}_{x_0}(r_0+r_1+\cdots+r_{N-1}+r_N\geq c) \\ \text{subject to} & \text{(i)}\ \ X_{n+1}\sim p(\cdot\,|x_n,u_n) \\ & \text{(ii)}\ \ u_n\in U \end{array} \qquad n=0,1,\ldots,N-1 \tag{1}$$

ただし $P^{\sigma}_{x_0}$ は、初期状態 x_0, マルコフ推移法則 p および 一般政策 σ から履歴の直積空間

$$H_N = X\times U\times X\times U\cdots\times U\times X \quad (2N+1)\text{個}$$

上に唯一定まる確率測度である。また、この確率測度による期待値作用素を $E^{\sigma}_{x_0}$ で表す。

3　一般政策クラス問題

この節では、次のような決定列の選択方法を考える。すなわち、与えられた初期（第0段）状態 x_0 を観て、まず決定 $u_0\in U$ を選択する。次に、マルコフ推移法則 $p(\cdot|x_0,u_0)$ に従って確率的に第1段状態 x_1 が出現する。このとき、これまでの二つ状態 x_0,x_1 に依存して第1段の決定 u_1 を取る。第2段には確率 $p(x_2|x_1,u_1)$ で状態 x_2 になる。このように決定列を選んでいって、第 $(N-1)$ 段には確率 $p(x_{N-1}|x_{N-2},u_{N-2})$ で状態 x_{N-1} になったとき、これまでの状態列 $x_0,\ x_1,\ \ldots,\ x_{N-1}$ を考慮して決定 u_{N-1} を選び、最終的に推移法則 $p(\cdot|x_{N-1},u_{N-1})$ に従って状態 x_N が現れる。このとき、第0段での決定の選択方法は関数 $\sigma_0: X\to U$ で指示される。第1段での選択方法は関数 $\sigma_1: X\times X\to U$ で指示され、一般に、第 n 段では関数 $\sigma_n: X\times X\times\cdots\times X\to U$ で示される。関数 σ_n を**第 n 一般決定関数**という。一般決定関数列

$$\sigma=\{\sigma_0,\sigma_1,\ldots,\sigma_{N-1}\}$$

を**一般政策**という。一般政策 σ は、それまでの状態列の出現に応じて意志決定者の決定を指定する。すなわち、第 n 段までに状態列 $x_0,\ x_1,\ \ldots,\ x_n$ を経てきたとき、意志決定者は決定 $\sigma_n(x_0,x_1,\ldots,x_n)\in U$ を取ることを表している。

したがって、意志決定者が一般政策 σ を採用すると、最大化問題 (1) の閾値確率は「部分」多重和

$$\begin{aligned} &P^{\sigma}_{x_0}(r_0+r_1+\cdots+r_{N-1}+r_N\geq c) \\ &=\sum_{(x_1,x_2,\ldots,x_N)\in(*)}\sum\cdots\sum p(x_1|x_0,u_0)p(x_2|x_1,u_1)\cdots p(x_N|x_{N-1},u_{N-1}) \end{aligned} \tag{2}$$

で表される。ただし、多重和をとる領域 $(*)$ は

$$r_0(x_0,u_0)+r_1(x_1,u_1)+\cdots+r_{N-1}(x_{N-1},u_{N-1})+r_N(x_N)\ \ \geq\ \ c \tag{3}$$

を満たす $(x_1,x_2,\ldots,x_N)\in X\times X\times\cdots\times X$ 全体にわたる多重和である。ここに、式(2),(3)における決定列 $\{u_0,\ u_1,\ \cdots,\ u_{N-1}\}$ は一般政策 $\sigma=\{\sigma_0,\ldots,\sigma_{N-1}\}$ を通して定まっていることに注意すべきである：

$$u_0=\sigma_0(x_0),\ u_1=\sigma_0(x_0,x_1),\ \ldots,\ u_{N-1}=\sigma_0(x_0,x_1,\ldots,x_{N-1}).$$

原問題 (1) では、逐次制約条件 (i), (ii) $0 \leq n \leq N-1$ を明示して、一般政策 σ を選ぶ問題としてダイナミックに表現している。しかし、実質的に同じことであるが、一般政策全体 $\Pi(g)$ 上の最大化としてスタティックに

$$\mathrm{P}_0(x_0) \qquad \text{Maximize} \quad P^{\sigma}_{x_0}(r_0 + \cdots + r_{N-1} + r_N \geq c) \quad \text{subject to} \quad \sigma \in \Pi(g)$$

で表すこともできることに注意しよう。

さて、われわれの求める最適解は問題 (1) の最大値関数 $v_0 = v_0(x_0)$ および最大値を与える最適政策 σ^* である：

$$\begin{aligned} v_0(x_0) &= P^{\sigma^*}_{x_0}(r_0 + \cdots + r_{N-1} + r_N \geq c) \\ &= \operatorname*{Max}_{\sigma \in \Pi(g)} P^{\sigma}_{x_0}(r_0 + \cdots + r_{N-1} + r_N \geq c) \quad x_0 \in X. \end{aligned}$$

この問題を**一般政策クラス問題**、または短く**一般問題**と呼ぶ。

次節では一般問題を等価な問題に変換して上記の最適解を求める。そのため、閾値確率を期待値に変換しておこう。

一般に、確率変数 Y が c 以上になる確率 $P(Y \geq c)$ は、数直線 R^1 上の区間 $[c, \infty)$ の定義関数

$$\psi(y) := 1_{[c,\infty)}(y) := \begin{cases} 1 & y \geq c \\ 0 & \text{その他} \end{cases}$$

を通した確率変数 $\psi(Y)$ の期待値 $E[\psi(Y)]$ で表される：

$$P(Y \geq c) = E[\psi(Y)].$$

このことに注意すると、一般問題 (1) の閾値確率は定義関数 $\psi = \psi(y)$ を通した期待値になる：

$$P^{\sigma}_{x_0}(r_0 + \cdots + r_{N-1} + r_N \geq c) \;=\; E^{\sigma}_{x_0}[\psi(r_0 + \cdots + r_{N-1} + r_N)].$$

すなわち、「部分」多重和は定義関数を通した「全」多重和に等しい：

$$\begin{aligned} &\sum_{(x_1,x_2,\ldots,x_N)\in(*)} \sum \cdots \sum p(x_1|x_0,u_0)p(x_2|x_1,u_1)\cdots p(x_N|x_{N-1},u_{N-1}) \\ &= \sum_{(x_1,x_2,\ldots,x_N)\in X\times X\times\cdots\times X} \sum \cdots \sum \{\psi(r_0(x_0,u_0) + \cdots + r_{N-1}(x_{N-1},u_{N-1}) + r_N(x_N)) \\ &\qquad \times p(x_1|x_0,u_0)p(x_2|x_1,u_1)\cdots p(x_N|x_{N-1},u_{N-1})\}. \end{aligned}$$

4 拡大マルコフ政策クラス問題 I

前節の議論により、「閾値確率」最大化問題は次の「期待値」最大化問題になる：

$$\begin{aligned} &\text{Maximize} \quad E^{\sigma}_{x_0}[\psi(r_0 + r_1 + \cdots + r_{N-1} + r_N)] \\ &\text{subject to} \quad \begin{matrix} (\mathrm{i})_n & X_{n+1} \sim p(\cdot\,|x_n,u_n) \\ (\mathrm{ii})_n & u_n \in U \end{matrix} \qquad n = 0,1,\ldots,N-1 \end{aligned}$$

これはダイナミックに表されているが、スタティックには

$$\text{Maximize} \quad E^{\sigma}_{x_0}[\psi(r_0 + r_1 + \cdots + r_{N-1} + r_N)] \quad \text{subject to} \quad \sigma \in \Pi(g)$$

で表される。

この問題を、新たに過去値をパラメータとする問題に埋め込んで考える。まず、第 n 段までの**集積値**確率変数列 $\{\tilde{\Lambda}_n\}$ および それらが取り得る**過去値集合**列 $\{\Lambda_n\}$ をそれぞれ次で定義する [6–8, 11–13] :

$$\begin{aligned}\tilde{\Lambda}_0 &\triangleq 0\\ \tilde{\Lambda}_n &\triangleq r_0(X_0,U_0)+\cdots+r_{n-1}(X_{n-1},U_{n-1}) \quad n=1,\ldots,N\end{aligned} \tag{4}$$

$$\begin{aligned}\Lambda_0 &\triangleq \{0\}\\ \Lambda_n &\triangleq \{\lambda_n \,|\, \lambda_n = r_0(x_0,u_0)+\cdots+r_{n-1}(x_{n-1},u_{n-1}),\\ &\qquad (x_0,u_0,\ldots,x_{n-1},u_{n-1})\in X\times U\times\cdots\times X\times U\}\\ &\qquad n=1,\ldots,N\end{aligned}$$

このとき、総利得は次になる :

$$r_0+r_1+\cdots+r_{N-1}+r_N=\tilde{\Lambda}_N+r_N.$$

式 (4) は漸化式

$$\begin{aligned}\tilde{\Lambda}_0 &= 0\\ \tilde{\Lambda}_{n+1} &= \tilde{\Lambda}_n+r_n(X_n,U_n) \quad n=0,\ldots,N-1\end{aligned}$$

に同値である。また、相隣る過去値集合 $\{\Lambda_{n-1},\ \Lambda_n\}$ 間には次の前向き再帰式が成り立つ :

補題 4.1

$$\begin{aligned}&\Lambda_0=\{0\}\\ &\Lambda_n=\{\lambda+r_{n-1}(x,u)\,|\,\lambda\in\Lambda_{n-1},\ (x,u)\in X\times U\} \quad n=1,2,\ldots,N.\end{aligned}$$

さらに、本来の状態空間 X に過去値集合を貼り合せた拡大状態空間列 $\{Y_n\}$ を直積で定義する :

$$Y_n \triangleq X\times\Lambda_n \qquad n=0,1,\ldots,N.$$

この新状態空間列上のマルコフ政策 $\gamma=\{\gamma_0,\gamma_1,\ldots,\gamma_{N-1}\}$ は**マルコフ決定関数列**

$$\gamma_n : Y_n\to U,\ \ (n=1,\ 2,\ \ldots,N)$$

で定まる。これを過去値による**拡大マルコフ政策**といい、その全体を $\tilde{\Pi}$ で表す。新たに**終端利得関数** T を

$$T(x;\lambda)\triangleq\psi(\lambda+r_N(x)) \quad (x;\lambda)\in Y_N$$

で定義する。さらに、定常マルコフ推移法則 $p=\{p(y|x,u)\}$ および パラメータ・ダイナミックス $\{\lambda_{n+1}=\lambda_n+r_n(x_n,u_n)\}$ によって拡大状態空間列上に定まる**非定常マルコフ推移法則** $q=\{q_n\}$ を

$$q_n((y;\mu)\,|\,(x;\lambda),\,u)\triangleq\begin{cases}p(y|x,\,u) & \lambda+r_{n-1}(x,u)=\mu\ \text{のとき}\\ 0 & \text{その他.}\end{cases}$$

で定義する。このとき、拡大マルコフ政策空間上の**終端型評価問題**

$$\mathrm{Q}_0(y_0)\qquad \begin{array}{ll} \text{Maximize} & \tilde{E}^{\gamma}_{y_0}[\psi(\tilde{\Lambda}_N + r_N(X_N))] \\ \text{subject to} & (\mathrm{i})_n \quad X_{n+1} \sim p(\cdot\,|x_n, u_n) \\ & (\mathrm{i})'_n \quad \tilde{\Lambda}_{n+1} = \tilde{\Lambda}_n + r_n(X_n, U_n) \quad n = 0, 1, \ldots, N-1 \\ & (\mathrm{ii})_n \quad u_n \in U \end{array}$$

を考える。ただし、 $y_0 = (x_0; 0)$. ここに $\tilde{E}^{\gamma}_{y_0}$ は、初期状態 y_0 、拡大マルコフ政策 γ および新マルコフ推移法則 q によって拡大状態空間列上に定まる確率測度 $\tilde{P}^{\gamma}_{y_0}$ に基づく期待値作用素である ([10])。

さて、第 n 段の状態 $y_n = (x_n; \lambda_n)(\in Y_n)$ から始まる部分過程

$$\mathrm{Q}_n(y_n)\qquad \begin{array}{ll} \text{Maximize} & \tilde{E}^{\gamma}_{y_n}[\psi(\tilde{\Lambda}_N + r_N(X_N))] \\ \text{subject to} & (\mathrm{i})_n \quad X_{m+1} \sim p(\cdot\,|x_m, u_m) \\ & (\mathrm{i})'_n \quad \tilde{\Lambda}_{m+1} = \tilde{\Lambda}_m + r_m(X_m, U_m) \quad m = n, \ldots, N-1 \\ & (\mathrm{ii})_n \quad u_m \in U \end{array}$$

の最大値を $u^n(x_n; \lambda_n)$ とする。ただし

$$u^N(x_N; \lambda_N) \triangleq \psi(\lambda_N + r_N(x_N)) \qquad (x_N; \lambda_N) \in Y_N.$$

このとき、次の後向きの再帰式が成り立つ：

定理 4.1

$$\begin{aligned} u^N(x; \lambda) &= \psi(\lambda + r_N(x)) \qquad x \in X,\ \lambda \in \Lambda_N \\ u^n(x; \lambda) &= \max_{u \in U} \sum_{y \in X} u^{n+1}(y; \lambda + r_n(x, u)) p(y|x, u) \qquad (5) \\ &\qquad\qquad x \in X,\ \lambda \in \Lambda_n,\ \ 0 \le n \le N-1. \end{aligned}$$

さて、式 (5) の最大（値を与える）点の全体を $\gamma^*_n(x; \lambda)$ とすると、拡大マルコフ政策クラス $\tilde{\Pi}$ の中での最適政策 $\gamma^* = \{\gamma^*_0, \gamma^*_1, \ldots, \gamma^*_{N-1}\}$ が得られる ([10, Theorem 4.2])。さらに、γ^* から、以下のように一般政策 $\sigma^* = \{\sigma^*_0, \sigma^*_1, \ldots, \sigma^*_{N-1}\}$ を生成する。すなわち、 $\sigma^*_n(x_0, x_1, \ldots, x_n)$ は

$$\begin{aligned} & u_0 := \gamma^*_1(x_0; 0), \quad \lambda_1 := 0 + r(x_0, u_0) \\ & u_1 := \gamma^*_1(x_1; \lambda_1), \quad \lambda_2 := \lambda_1 + r(x_1, u_1) \\ & \qquad\vdots \\ & u_{n-1} := \gamma^*_{n-1}(x_{n-1}; \lambda_{n-1}), \quad \lambda_n := \lambda_{n-1} + r(x_{n-1}, u_{n-1}) \\ & \sigma^*_n(x_0, x_1, \ldots, x_n) := \gamma^*_n(x_n; \lambda_n). \end{aligned}$$

このとき、次が成り立つ：

定理 4.2 *([10, Theorem 6.1])*
(i) 政策 σ^* は一般政策クラス $\Pi(g)$ の中での最適である。
(ii) 拡大マルコフ政策クラス $\tilde{\Pi}$ の最大値は一般政策クラス $\Pi(g)$ の最大値に等しい：

$$u^0(x_0; 0) = v_0(x_0).$$

5 拡大マルコフクラス問題 II

この節では、与えられた水準値 c を固定して、もう1つの拡大マルコフ政策クラスの中で最適化しても最適政策が得られることを示す。このマルコフクラス上の閾値確率最大化問題は

$$\mathrm{M}_0(x_0)\quad \begin{array}{ll} \text{Maximize} & P^{\tau}_{x_0}(r_0+r_1+\cdots+r_{N-1}+r_N \geq c) \\ \text{subject to} & \text{(i)}_n \quad X_{n+1}\sim p(\cdot|x_n,u_n) \\ & \text{(ii)}_n \quad u_n \in U \end{array} \qquad n=0,1,\ldots,N-1 \tag{6}$$

で表される。この問題では、マルコフ的に推移する本来の状態変数 X_n に加えて、新たに導入する将来の閾値 c_n が所与の閾値 $c_0=c$ から時間と共に確定的に変化すると、捉える。すなわち、対 $(X_n;c_n)$ を新状態変数とし、その新たなマルコフ推移のもとで所与の閾値確率を最小化する。そのため、まずここで、**未来閾値集合列** $\{C_n\}$ を

$$\begin{aligned} C_0 &\triangleq \{c_0 \,|\, c_0=c\} \\ C_n &\triangleq \left\{ c_n \;\middle|\; \begin{array}{c} c_n = c - r_0(x_0,u_0) - \cdots - r_{n-1}(x_{n-1},u_{n-1}) \\ (x_0,u_0,\ldots,x_{n-1},u_{n-1}) \in X\times U\times\cdots\times X\times U \end{array} \right\} \\ & \qquad\qquad n=1,\ldots,N \end{aligned}$$

で導入しておく。C_n は未来の時刻 n での可能な閾値の集合を表している。このとき、列 $\{C_n\}$ は次の前向きの再帰式を満たす：

補題 5.1

$$\begin{aligned} C_0 &= \{c\} \\ C_{n+1} &= \{c-r(x,u)\,|\,c\in C_n,\ (x,u)\in X\times U\} \\ & \qquad\qquad 0\leq n\leq N-1. \end{aligned}$$

したがって、時刻 n で「拡大」状態 $(x_n;c_n)$ において意思決定者が決定 u_n を選ぶと、次の時刻 $(n+1)$ では $(x_{n+1};c_{n+1})$ に確率 $p(x_{n+1}|x_n,u_n)$ で推移する。ただし、第2成分は確定的に定まる： $c_{n+1}:=c_n-r(x_n,u_n)$.

このとき、決定関数

$$\tau_n : X\times C_n \to U \quad 0\leq n\leq N-1$$

から成る列 $\tau=\{\tau_0,\ldots,\tau_{N-1}\}$ を未来（閾）値による**「拡大」マルコフ政策**という。この拡大マルコフ政策の全体を $\bar{\Pi}$ で表す。意思決定者が拡大マルコフ政策 $\tau(\in\bar{\Pi})$ を採用すると、最大化問題 (6) の閾値確率は「部分」多重和

$$\begin{aligned} & P^{\tau}_{x_0}(r_0+r_1+\cdots+r_{N-1}+r_N\geq c) \\ &= \sum_{(x_1,x_2,\ldots,x_N)\in(*)}\sum\cdots\sum p(x_1|x_0,u_0)p(x_2|x_1,u_1)\cdots p(x_N|x_{N-1},u_{N-1}) \end{aligned} \tag{7}$$

で表される。ただし、多重和をとる領域 $(*)$ は

$$r_0(x_0,u_0)+r_1(x_1,u_1)+\cdots+r_{N-1}(x_{N-1},u_{N-1})+r_N(x_N) \;\geq\; c \tag{8}$$

を満たす $(x_1, x_2, \ldots, x_N) \in X\times X\times\cdots\times X$ 全体にわたる多重和である。ここに、式(7),(8)における決定列 $\{u_0,\ u_1,\ \cdots,\ u_{N-1}\}$ はマルコフ政策 τ の決定関数列を通して定まっていることに注意すべきである：

$$u_0 = \tau_0(x_0; c_0),\ u_1 = \tau_1(x_1; c_1),\ \ldots,\ u_{N-1} = \tau_{N-1}(x_{N-1}; c_{N-1}).$$

このとき、次の関係式を得る。

補題 5.2 任意のマルコフ政策 $\tau = \{\tau_n, \tau_{n+1}, \ldots, \tau_{N-1}\}$ と任意の $(x_n; x_n) \in X\times C_n$ に対して、

$$\begin{aligned} &P^{\tau}_{x_n}(r_n + \cdots + r_N \geq c_n) \\ &= \sum_{x_{n+1}\in X} P^{\tau'}_{x_{n+1}}(r_{n+1} + \cdots + r_N \geq c_n - r_n)p(x_{n+1}|x_n, u_n) \end{aligned}$$

が成り立つ。ここに

$$r_n = r(x_n, u_n), \quad u_n = \tau_n(x_n; c_n), \quad \tau' = \{\tau_{n+1}, \ldots, \tau_{N-1}\}.$$

補題5.2を多重和で表すと、次になる。

補題 5.3 任意のマルコフ政策 $\tau = \{\tau_n, \tau_{n+1}, \ldots, \tau_{N-1}\}$ と任意の $x_n \in X$ に対して、

$$\begin{aligned} &\sum\sum\cdots\sum_{(x_{n+1}, x_{n+2}, \ldots, x_N)\in(*)} p_{n+1}p_{n+2}\cdots p_N \\ &= \sum_{x_{n+1}\in X}\left[\sum\sum\cdots\sum_{(x_{n+2}, \ldots, x_N)\in(\star)} p_{n+2}\cdots p_N\right]p(x_{n+1}|x_n, u_n) \end{aligned}$$

が成り立つ。ここに

$$p_{m+1} = p(x_{m+1}|x_m, u_m),\ u_m = \tau_m(x_m; c_m),\ c_{m+1} = c_m - r(x_m, c_m)\ \ n \leq m \leq N-1.$$

また、$(*)$ は $r_n(x_n, u_n) + \cdots + r_N(x_N) \geq c_n$ を満たす $(x_{n+1}, \ldots, x_N) \in X\times\cdots\times X$ 全体にわたる多重和であり、$(\star)$ は $r_{n+1}(x_{n+1}, u_{n+1}) + \cdots + r_N(x_N) \geq c_n - r_n$ を満たす $(x_{n+2}, \ldots, x_N)$ 全体にわたっている。ただし、$r_n = r(x_n, u_n),\ u_n = \tau_n(x_n; c_n)$.

したがって、上述の補題から後向きの再帰式が成り立つ：

定理 5.1

$$f_n(x; c) = \operatorname*{Max}_{u\in U}\sum_{y\in X} f_{n+1}(y; c - r(x, u))p(y|x, u) \qquad (9)$$
$$(x; c) \in X\times C_n,\ \ 0 \leq n \leq N-1$$

$$f_N(x; c) = \begin{cases} 1 & \text{if } r(x) \geq c \\ 0 & \text{otherwise} \end{cases} \qquad (x; c) \in X\times C_N.$$

さて、式 (9) の最大点の全体を $\bar{\tau}_n(x;c)$ とすると、拡大マルコフ政策クラス $\bar{\Pi}$ の中での最適政策 $\bar{\tau} = \{\bar{\tau}_0, \bar{\tau}_1, \ldots, \bar{\tau}_{N-1}\}$ が得られる。さらに、$\bar{\tau}$ から、以下のように一般政策 $\bar{\sigma} = \{\bar{\sigma}_0, \bar{\sigma}_1, \ldots, \bar{\sigma}_{N-1}\}$ を生成する。すなわち、$\bar{\sigma}_n(x_0, x_1, \ldots, x_n)$ は

$$
\begin{aligned}
&u_0 := \bar{\tau}_0(x_0;c), \quad c_1 := c - r(x_0, u_0)\\
&u_1 := \bar{\tau}_1(x_1;c_1), \quad c_2 := c_1 - r(x_1, u_1)\\
&\qquad\vdots\\
&u_{n-1} := \bar{\tau}_{n-1}(x_{n-1};c_{n-1}), \quad c_n := c_{n-1} - r(x_{n-1}, u_{n-1})\\
&\bar{\sigma}_n(x_0, x_1, \ldots, x_n) := \bar{\tau}_n(x_n;c_n).
\end{aligned}
$$

このとき、次が成り立つ：

定理 5.2 *([10, Theorem 6.1])*
(i) 政策 $\bar{\sigma}$ は一般政策クラス $\Pi(g)$ の中での最適である。
(ii) 拡大マルコフ政策クラス $\tilde{\Pi}$ の最大値は一般政策クラス $\Pi(g)$ の最大値に等しい：

$$f_0(x_0;c) = v_0(x_0).$$

さらに、過去値に基づく拡大マルコフクラス問題 I と未来（閾）値に基づく拡大マルコフクラス問題 II の間には次のような相補的な双対性が成り立っている。

定理 5.3 (相補的双対定理)
(i) 任意の $\lambda_n \in \Lambda_n$ に対してある $c_n \in C_n$ が存在して、その和は一定値 c である：

$$\lambda_n + c_n = c.$$

逆も成り立つ。すなわち、任意の $c_n \in C_n$ に対してある $\lambda_n \in \Lambda_n$ が存在してその和は一定値 c である。
(ii) このとき、最適解（最適値と最適決定関数）は共に一致している：

$$u^n(x_n;\lambda_n) = f_n(x_n;c_n), \quad \gamma_n^*(x_n;\lambda_n) = \bar{\tau}_n(x_n;c_n) \qquad x_n \in X.$$

(iii) 過去値に基づく拡大マルコフクラス問題の最適解と未来（閾）値に基づく拡大マルコフクラス問題の最適解は上述の定和という意味で一致している：

$$
\begin{aligned}
u^n(x_n;c-c_n) = f_n(x_n;c_n), \qquad & \gamma_n^*(x_n;c-c_n) = \bar{\tau}_n(x_n;c_n)\\
& (x_n;c_n) \in X{\times}C_n, \quad 0 \le n \le N.
\end{aligned}
$$

すなわち

$$
\begin{aligned}
f_n(x_n;c-\lambda_n) = u^n(x_n;\lambda_n), \qquad & \bar{\tau}_n(x_n;c-\lambda_n) = \gamma_n^*(x_n;\lambda_n)\\
& (x_n;\lambda_n) \in X{\times}\Lambda_n, \quad 0 \le n \le N.
\end{aligned}
$$

定理 5.4 （一致定理）
過去値による拡大最適政策 γ^* から生成された一般最適政策 σ^* は未来（閾）値による拡大最適政策 $\bar{\tau}$ から生成された一般最適政策 $\bar{\sigma}$ に一致している：

$$\sigma^* = \bar{\sigma}.$$

6　3状態2決定2段問題

この節では、3-2-2（3状態2決定2段）モデルにおいて（加法型）総合評価値が $c = 2.5$ 以上になる閾値確率を最大化する問題を考える：

$$\begin{aligned} \text{Maximize} \quad & P^{\sigma}_{x_0}(\, r_0(U_0) + r_1(U_1) + r_2(X_2) \geq 2.5\,) \\ \text{subject to} \quad & \text{(i)} \;\; X_{n+1} \sim p(\cdot|x_n, u_n) \quad n = 0, 1 \\ & \text{(ii)} \;\; u_0 \in U,\; u_1 \in U \end{aligned}$$

ただし、データとしては、次のような Bellman and Zadeh [1, pp.B154] の数値例を用いる：

$$r_2(s_1) = 0.3 \qquad r_2(s_2) = 1.0 \qquad r_2(s_3) = 0.8$$

$$r_1(a_1) = 1.0 \qquad r_1(a_2) = 0.6$$

$$r_0(a_1) = 0.7 \qquad r_0(a_2) = 1.0$$

$u_t = a_1$

$x_t \setminus x_{t+1}$	s_1	s_2	s_3
s_1	*0.8*	*0.1*	*0.1*
s_2	*0.0*	*0.1*	*0.9*
s_3	*0.8*	*0.1*	*0.1*

$u_t = a_2$

$x_t \setminus x_{t+1}$	s_1	s_2	s_3
s_1	*0.1*	*0.9*	*0.0*
s_2	*0.8*	*0.1*	*0.1*
s_3	*0.1*	*0.0*	*0.9*

6.1　拡大マルコフ政策 I

最初に、定義

$$\begin{aligned} \Lambda_0 &= \{0\} \\ \Lambda_1 &= \{\lambda_1 \,|\, \lambda_1 = r_0(u_0),\; u_0 \in U\} \\ \Lambda_2 &= \{\lambda_2 \,|\, \lambda_2 = r_0(u_0) + r_1(u_1),\; u_0, u_1 \in U\} \end{aligned}$$

より、過去値集合列

$$\Lambda_0 = \{0\}, \quad \Lambda_1 = \{0.7, 1.0\}, \quad \Lambda_2 = \{1.3, 1.6, 1.7, 2.0\}$$

を求めておく。このとき、集積値確率変数

$$\tilde{\Lambda}_0 = 0, \quad \tilde{\Lambda}_1 = r_0(U_0), \quad \tilde{\Lambda}_2 = r_0(U_0) + r_1(U_1)$$

を用いると、拡大状態空間上の終端問題は

$$\begin{aligned} \text{Maximize} \quad & \tilde{E}^{\gamma}_{y_0}[\psi(\tilde{\Lambda}_2 + r_2(X_2)] \qquad (y_0 = (x_0; 0)) \\ \text{subject to} \quad & \text{(i)} \;\; X_{n+1} \sim p(\cdot|x_n, u_n) \\ & \text{(i)}' \;\; \tilde{\Lambda}_{n+1} = \tilde{\Lambda}_n + r_n(U_n) \quad n = 0, 1 \\ & \text{(ii)} \;\; u_n \in \{a_1, a_2\} \end{aligned}$$

になる。ただし $\psi(y) = 1_{[2.5,\infty)}(y)$, $\gamma = \{\gamma_0, \gamma_1\}$.

まず、終端関数

$$u^2(x_2;\lambda_2) \;=\; \psi(\lambda_2 + r_2(x_2))$$

は次の表１の通り：

$x_2 \backslash \lambda_2$	1.3	1.6	1.7	2.0
s_1	0	0	0	0
s_2	0	1	1	1
s_3	0	0	1	1

表１ : $u^2(x_2;\lambda_2)$

次に、第２最適値関数

$$u^1(x_1;\lambda_1) \;=\; \underset{u_1}{\mathrm{Max}} \sum_{x_2} u^2(x_2;\lambda_1 + r_1(u_1))p(x_2|x_1,u_1)$$

を計算するのに表１を用いる。最初の $u^1(s_1;0.7)$ は

$$\begin{aligned}
u^1(s_1;0.7) &= [u^2(s_1;0.7+1.0)0.8 + u^2(s_2;0.7+1.0)0.1 + u^2(s_3;0.7+1.0)0.1] \\
&\quad \vee [u^2(s_1;0.7+0.6)0.1 + u^2(s_2;0.7+0.6)0.9 + u^2(s_3;0.7+0.6)0.0] \\
&= [u^2(s_1;1.7)0.8 + u^2(s_2;1.7)0.1 + u^2(s_3;1.7)0.1] \\
&\quad \vee [u^2(s_1;1.3)0.1 + u^2(s_2;1.3)0.9 + u^2(s_3;1.3)0.0] \\
&= [0\times 0.8 + 1\times 0.1 + 1\times 0.1] \vee [0\times 0.1 + 0\times 0.9 + 0\times 0.0] \\
&= 0.2 \vee 0 \\
&= 0.2 \quad \gamma_1^*(s_1;0.7) = a_1
\end{aligned}$$

になる。以下、同様にすると、第２最適値関数 u^1 と第２最適決定関数 γ_1^* は次の表２になる

$x_1 \backslash \lambda_1$	0.7		1.0	
s_1	0.2	a_1	0.9	a_2
s_2	1.0	a_1	1.0	a_1
s_3	0.2	a_1	0.2	a_1

表２ : $u^1(x_1;\lambda_1)$ $\gamma_1^*(x_1;\lambda_1)$

最後に、第１最適値関数

$$u^0(x_0;\lambda_0) \;=\; \underset{u_0}{\mathrm{Max}} \sum_{x_1} u^1(x_1;\lambda_0 + r_0(u_0))p(x_1|x_0,u_0)$$

を求める。 $u^0(s_1;0.7)$ は表２を用いると

$$\begin{aligned} u^0(s_1;0) &= [u^1(s_1;0.7)0.8+u^1(s_2;0.7)0.1+u^1(s_3;0.7)0.1] \\ &\quad \vee[u^1(s_1;1.0)0.1+u^1(s_2;1.0)0.9+u^1(s_3;1.0)0.0] \\ &= [0.2\times 0.8+1.0\times 0.1+0.2\times 0.1]\vee[0.9\times 0.1+1.0\times 0.9+0.2\times 0.0] \\ &= 0.28\vee 0.99 \\ &= 0.99 \quad \gamma_0^*(s_1;0)=a_2 \end{aligned}$$

になる。同様にすると、第１最適値関数 u^0 と第１最適決定関数 γ_0^* は次の表３になる：

x_0	$u^0(x_0;0)$	$\gamma_0^*(x_0;0)$
s_1	0.99	a_2
s_2	0.84	a_2
s_3	0.28	a_1

表３： $u^0(x_0;0)$ $\gamma_0^*(x_0;0)$

最適解をまとめると、次の表４になる。

	$u^2(x_2;\lambda_2)$				$u^1(x_1;\lambda_1)$	$\gamma_1^*(x_1;\lambda_1)$	$u^0(x_0;0)$	$\gamma_0^*(x_0;0)$
$x_n \diagdown \lambda_n$	1.3	1.6	1.7	2.0	0.7	1.0	0	
s_1	0	0	0	0	0.2	a_1	0.9 a_2	0.99 a_2
s_2	0	1	1	1	1.0	a_1	1.0 a_1	0.84 a_2
s_3	0	0	1	1	0.2	a_1	0.2 a_1	0.28 a_1

表４： 拡大マルコフ政策クラスＩの最適解

さて、埋没問題の最適（拡大マルコフ）政策 $\gamma^*=\{\gamma_0^*,\gamma_1^*\}$ から、式

$$\begin{aligned} \sigma_0^*(x_0) &:= \gamma_0^*(x_0;0) \\ & \quad u_0:=\gamma_0^*(x_0;0), \quad \lambda_1:=r_0(u_0) \\ \sigma_0^*(x_0,x_1) &:= \gamma_1^*(x_1;\lambda_1) \end{aligned}$$

によって、最適（一般）政策 $\sigma^*=\{\sigma_0^*,\sigma_1^*\}$ を構成しよう。まず、第１決定関数

$$\begin{aligned} \sigma_0^*(s_1)&=\gamma_0^*(s_1;0)=a_2 \\ \sigma_0^*(s_2)&=\gamma_0^*(s_2;0)=a_2 \\ \sigma_0^*(s_3)&=\gamma_0^*(s_3;0)=a_1 \end{aligned}$$

が得られる。次に、第２決定関数は以下になる：

$$\begin{aligned} \sigma_1^*(s_1,s_1)&=\gamma_1^*(s_1;1.0)=a_2 \\ \sigma_1^*(s_1,s_2)&=\gamma_1^*(s_2;1.0)=a_1 \\ \sigma_1^*(s_1,s_3)&=\gamma_1^*(s_3;1.0)=a_1 \end{aligned}$$

$$\sigma_1^*(s_2, s_1) = \gamma_1^*(s_1; 1.0) = a_2$$
$$\sigma_1^*(s_2, s_2) = \gamma_1^*(s_2; 1.0) = a_1$$
$$\sigma_1^*(s_2, s_3) = \gamma_1^*(s_3; 1.0) = a_1$$

$$\sigma_1^*(s_3, s_1) = \gamma_1^*(s_1; 0.7) = a_1$$
$$\sigma_1^*(s_3, s_2) = \gamma_1^*(s_2; 0.7) = a_1$$
$$\sigma_1^*(s_3, s_3) = \gamma_1^*(s_3; 0.7) = a_1$$

すなわち、不変埋没原理によって（マルコフでない!!）最適政策 σ^* が得られた。

6.2 拡大マルコフ政策 II

最初に、定義

$$\begin{aligned} C_0 &= \{2.5\} \\ C_1 &= \{c_1 \,|\, c_1 = 2.5 - r_0(u_0),\ u_0 \in U\} \\ C_2 &= \{c_2 \,|\, c_2 = 2.5 - r_0(u_0) - r_1(u_1),\ u_0, u_1 \in U\} \end{aligned}$$

より、未来（閾）値集合列

$$C_0 = \{2.5\}, \quad C_1 = \{1.8, 1.5\}, \quad C_2 = \{1.2, 0.9, 0.8, 0.5\}$$

を求めておく。

まず、 f_2 を

$$f_2(x_2; c_2) = \begin{cases} 1 & \text{if } r(x_2) \geq c_2 \\ 0 & \text{otherwise} \end{cases} \qquad (x_2; c_2) \in X \times C_2$$

で計算すると、は次の表５が得られる：

$x_2 \diagdown c_2$	1.2	0.9	0.6	0.5
s_1	0	0	0	0
s_2	0	1	1	1
s_3	0	0	1	1

表５ : $f_2(x_2; c_2)$

次に、 表５を用いて、f_1 を

$$f_1(x_1; c_1) = \operatorname*{Max}_{u_1} \sum_{x_2} f_2(x_2; c_1 - r_1(u_1)) p(x_2 | x_1, u_1)$$

で計算する。最初の $f_1(s_1;1.8)$ は

$$\begin{aligned} f_1(s_1;1.8) &= [f_2(s_1;1.8-1.0)0.8+f_2(s_2;1.8-1.0)0.1+f_2(s_3;1.8-1.0)0.1] \\ &\quad \vee[f_2(s_1;1.8-0.6)0.1+f_2(s_2;1.8-0.6)0.9+f_2(s_3;1.8-0.6)0.0] \\ &= [f_2(s_1;0.8)0.8+f_2(s_2;0.8)0.1+f_2(s_3;0.8)0.1] \\ &\quad \vee[f_2(s_1;1.2)0.1+f_2(s_2;1.2)0.9+f_2(s_3;1.2)0.0] \\ &= [0\times 0.8+1\times 0.1+1\times 0.1]\vee[0\times 0.1+0\times 0.9+0\times 0.0] \\ &= 0.2\vee 0 \\ &= 0.2 \quad \bar{\tau}_1(s_1;0.7)=a_1 \end{aligned}$$

になる。以下、同様にすると、第２最適値関数 f_1 と第２最適決定関数 $\bar{\tau}_1$ は次の表６になる

$x_1 \diagdown c_1$	1.8		1.5	
s_1	0.2	a_1	0.9	a_2
s_2	1.0	a_1	1.0	a_1
s_3	0.2	a_1	0.2	a_1

表６： $f_1(x_1;c_1)$ $\bar{\tau}_1(x_1;c_1)$

最後に、第１最適値関数

$$f_0(x_0;c_0) = \underset{u_0}{\mathrm{Max}} \sum_{x_1} f_1(x_1;c_0-r_0(u_0))p(x_1|x_0,u_0)$$

を求める。 $f_0(s_1;2.5)$ は表６を用いると

$$\begin{aligned} f_0(s_1;2.5) &= [f_1(s_1;1.8)0.8+f_1(s_2;1.8)0.1+f_1(s_3;1.8)0.1] \\ &\quad \vee[f_1(s_1;1.5)0.1+f_1(s_2;1.5)0.9+f_1(s_3;1.5)0.0] \\ &= [0.2\times 0.8+1.0\times 0.1+0.2\times 0.1]\vee[0.9\times 0.1+1.0\times 0.9+0.2\times 0.0] \\ &= 0.28\vee 0.99 \\ &= 0.99 \quad \bar{\tau}_0(s_1;0)=a_2 \end{aligned}$$

になる。同様にすると、第１最適値関数 f_0 と第１最適決定関数 $\bar{\tau}_0$ は次の表７になる：

x_0	$f_0(x_0;2.5)$	$\bar{\tau}_0(x_0;2.5)$
s_1	0.99	a_2
s_2	0.84	a_2
s_3	0.28	a_1

表７： $f_0(x_0;2.5)$ $\bar{\tau}_0(x_0;2.5)$

最適解をまとめると、次の表８になる。

	$f_2(x_2;c_2)$				$f_1(x_1;c_1)$	$\bar{\tau}_1(x_1;c_1)$	$f_0(x_0;2.5)$	$\bar{\tau}_0(x_0;2.5)$
$x_n \diagdown \lambda_n$	1.2	0.9	0.8	0.5	1.8		1.5	2.5
s_1	0	0	0	0	0.2 a_1		0.9 a_2	0.99 a_2
s_2	0	1	1	1	1.0 a_1		1.0 a_1	0.84 a_2
s_3	0	0	1	1	0.2 a_1		0.2 a_1	0.28 a_1

表８： 拡大マルコフ政策クラスＩの最適解

さて、この埋没問題の最適（拡大マルコフ）政策 $\bar{\tau} = \{\bar{\tau}_0, \bar{\tau}_1\}$ から、式

$$\begin{aligned} \bar{\sigma}_0(x_0) &:= \bar{\tau}_0(x_0;2.5) \\ & \quad u_0 := \bar{\tau}_0(x_0;2.5), \quad c_1 := 2.5 - r_0(u_0) \\ \bar{\sigma}_0(x_0,x_1) &:= \bar{\tau}_1(x_1;c_1) \end{aligned}$$

によって、最適（一般）政策 $\bar{\sigma} = \{\bar{\sigma}_0, \bar{\sigma}_1\}$ を構成しよう。まず、第１決定関数

$$\begin{aligned} \bar{\sigma}_0(s_1) &= \bar{\tau}_0(s_1;2.5) = a_2 \\ \bar{\sigma}_0(s_2) &= \bar{\tau}_0(s_2;2.5) = a_2 \\ \bar{\sigma}_0(s_3) &= \bar{\tau}_0(s_3;2.5) = a_1 \end{aligned}$$

が得られる。次に、第２決定関数は以下になる：

$$\begin{aligned} \bar{\sigma}_1(s_1,s_1) &= \bar{\tau}_1(s_1;1.5) = a_2 \\ \bar{\sigma}_1(s_1,s_2) &= \bar{\tau}_1(s_2;1.5) = a_1 \\ \bar{\sigma}_1(s_1,s_3) &= \bar{\tau}_1(s_3;1.5) = a_1 \end{aligned}$$

$$\begin{aligned} \bar{\sigma}_1(s_2,s_1) &= \bar{\tau}_1(s_1;1.5) = a_2 \\ \bar{\sigma}_1(s_2,s_2) &= \bar{\tau}_1(s_2;1.5) = a_1 \\ \bar{\sigma}_1(s_2,s_3) &= \bar{\tau}_1(s_3;1.5) = a_1 \end{aligned}$$

$$\begin{aligned} \bar{\sigma}_1(s_3,s_1) &= \bar{\tau}_1(s_1;1.8) = a_1 \\ \bar{\sigma}_1(s_3,s_2) &= \bar{\tau}_1(s_2;1.8) = a_1 \\ \bar{\sigma}_1(s_3,s_3) &= \bar{\tau}_1(s_3;1.8) = a_1 \end{aligned}$$

すなわち、不変埋没原理によって（マルコフでない!!）最適政策 $\bar{\sigma}$ が得られた。

過去値による拡大最適政策 γ^* から生成された一般最適政策 σ^* は未来（閾）値による拡大最適政策 $\bar{\tau}$ から生成された一般最適政策 $\bar{\sigma}$ に一致している：

$$\sigma^* = \bar{\sigma}.$$

6.3 多段確率決定樹表

多段確率決定樹表は、いわゆる決定樹（ディシジョン・ツリー）、決定表（ディシジョン・テーブル）をそれぞれ進化・発展させ、多段階にわたる確率的決定過程の問題記述から最適解構成に至るまでを1枚に統合した図表である。問題のデータを過程の進行状況に応じて配列し、あらゆる可能な経路とその評価値・確率を図示し、各段における最適決定の選択を明示している。この意味では列挙法の解構成を与えている。しかし、最適解に至るまでは動的計画法の再帰式を解く順に構成されている。この樹表ではあらゆる型の評価関数に対してその期待値最適化が解かれる [5]。

まず、原始政策 $\mu = \{\mu_0, \mu_1\}$ に対して、条件付き期待値作用素 $E_{x_0}^{\mu_0}$, $E_{h_1}^{\mu_1}$ を次で定義する：

$$
\begin{aligned}
E_{x_0}^{\mu_0}[g_1] &\triangleq \sum_{x_1 \in X} g_1(x_0, u_0, x_1) p(x_1 | x_0, u_0) \\
&\quad \text{ただし } u_0 = \mu_0(x_0), \quad g_1 : X \times U \times X \to R^1 \\
E_{h_1}^{\mu_1}[g_2] &\triangleq \sum_{x_2 \in X} g_2(x_0, u_0, x_1, u_1, x_2) p(x_2 | x_1, u_1) \\
&\quad \text{ただし } u_1 = \mu_1(h_1);\ h_1 = (x_0, u_0, x_1), \\
&\qquad\qquad g_2 : X \times U \times X \times U \times X \to R^1.
\end{aligned}
$$

このとき、閾値確率は期待値になり、条件付き期待値作用素の繰り返しで表される：

$$
\begin{aligned}
& P_{x_0}^{\mu}(r_0 + r_1 + r_2 \geq c) \\
= \ & E_{x_0}^{\mu}[\psi(r_0 + r_1 + r_2)] \\
= \ & E_{x_0}^{\mu_0} E_{h_1}^{\mu_1}[\psi(r_0 + r_1 + r_2)]
\end{aligned}
$$

したがって、原始政策クラス $\Pi(p)$ 上の「同時」最適化は条件付き期待値の「繰り返し」最適化に等しい [3]：

$$
\begin{aligned}
& \operatorname*{Max}_{\mu \in \Pi(p)} P_{x_0}^{\mu}(r_0 + r_1 + r_2 \geq c) \\
= \ & \operatorname*{Max}_{\mu \in \Pi(p)} E_{x_0}^{\mu}[\psi(r_0 + r_1 + r_2)] \\
= \ & \operatorname*{Max}_{\mu_0} \operatorname*{Max}_{\mu_1} E_{x_0}^{\mu_0} E_{h_1}^{\mu_1}[\psi(r_0 + r_1 + r_2)] \\
= \ & \operatorname*{Max}_{\mu_0} E_{x_0}^{\mu_0} \operatorname*{Max}_{\mu_1} E_{h_1}^{\mu_1}[\psi(r_0 + r_1 + r_2)].
\end{aligned}
$$

これは動的計画法の原始的な型 (a primitive form) である [4]。すなわち、原始政策クラス問題の再帰式

$$
\begin{aligned}
w_2(x_0, u_0, x_1, u_1, x_2) &= \psi(r_0(u_0) + r_1(u_1) + r_2(x_2)) \\
w_1(x_0, u_0, x_1) &= \operatorname*{Max}_{u_1} \sum_{x_2} w_2(x_0, u_0, x_1, u_1, x_2) p(x_2 | x_1, u_1) \\
w_0(x_0) &= \operatorname*{Max}_{u_0} \sum_{x_0} w_1(x_0, u_0, x_1) p(x_1 | x_0, u_0)
\end{aligned}
$$

が得られたことになる。

図 1,2 および 3 では、次のように簡略化している：

$$\begin{aligned}
&\text{履歴} = x_0\ \ r_0(u_0)/u_0\ \ p_0\ \ x_1\ \ r_1(u_1)/u_1\ \ p_1\ \ x_2\ \ r_2(x_2)\\
&\qquad \text{ただし } p_0 = p(x_1 \mid x_0, u_0), \quad p_1 = p(x_2 \mid x_1, u_1)\\
&\text{加法} = \text{加法型評価値} = r_0(u_0) + r_1(u_1) + r_2(x_2)\\
&\text{経路} = \text{経路確率} = p_0 p_1\\
&\text{閾確} = \text{閾値確率} = \psi(r_0(u_0) + r_1(u_1) + r_2(x_2))p_0 p_1\\
&\qquad \text{ただし } \psi(y) = 1_{[2.5,\infty)}(y)\\
&\text{部分確} = \text{部分確率} = \sum_{x_2} \psi(r_0(u_0) + r_1(u_1) + r_2(x_2))p_0 p_1\\
&\text{全確率} = \text{全体確率} = \sum_{x_1}\sum_{x_2} \psi(r_0(u_0) + r_1(u_1) + r_2(x_2))p_0 p_1.
\end{aligned}$$

イタリック体は確率を表し、ボールド体は上下の確率のうち大きい方を選択したことを表す。特に、履歴の欄では５つの数値 $r_0 = r_0(u_0)$, p_0, $r_1 = r_1(u_1)$, p_1, $r_2 = r_2(x_2)$ のみを記している。

図１によって s_1 からの最適解（最大値および最適決定関数列）が求められる。まず、「部分確」の列は、過程が $x_0 = s_1 \to u_0 \to x_1$ まで進行してきたとき、各選択において（上下の）大きい値（= ゴチック体の数字) を原始最適決定

$$\tilde{\mu}_1(s_1, a_1, s_1) = a_1, \quad \tilde{\mu}_1(s_1, a_1, s_2) = a_1, \quad \tilde{\mu}_1(s_1, a_1, s_3) = a_1$$

$$\tilde{\mu}_1(s_1, a_2, s_1) = a_2, \quad \tilde{\mu}_1(s_1, a_2, s_2) = a_1, \quad \tilde{\mu}_1(s_1, a_2, s_3) = a_1, a_2$$

として選択していることを示している。すなわち、「部分確」の列のボールド体は、各 $h_1 = (s_1, u_0, x_1)$ に対して、最初の最適化

$$\underset{\mu_1}{\mathrm{Max}}\, E_{h_1}^{\mu_1}[\psi(r_0 + r_1 + r_2)]$$

を行なっている。この最大値を $g_1 = g_1(h_1)$ とする。

次に、「全確率」の列は、過程が $x_0 = s_1$ から出発すると、上下の比較によって $\tilde{\mu}_0(s_1) = a_2$ が最適決定で、最大値は $w_0(s_1) = \mathbf{0.99}$ あることを示している。したがって、一般最適決定

$$\tilde{\sigma}_0(s_1) = a_2; \qquad \tilde{\sigma}_1(s_1, s_1) = a_2, \quad \tilde{\sigma}_1(s_1, s_2) = a_1, \quad \tilde{\sigma}_1(s_1, s_3) = a_1, a_2$$

が得られる。すなわち、「全確率」の列のボールド体は $x_0 = s_1$ に対して次の最適化

$$\underset{\mu_0}{\mathrm{Max}}\, E_{x_0}^{\mu_0}[g_1]$$

を行なっている。

さらに、s_2, s_3 からの図 2,3 と合せると、原始最適政策 $\tilde{\mu} = \{\tilde{\mu}_0, \tilde{\mu}_1\}$ が得られ、最適解は

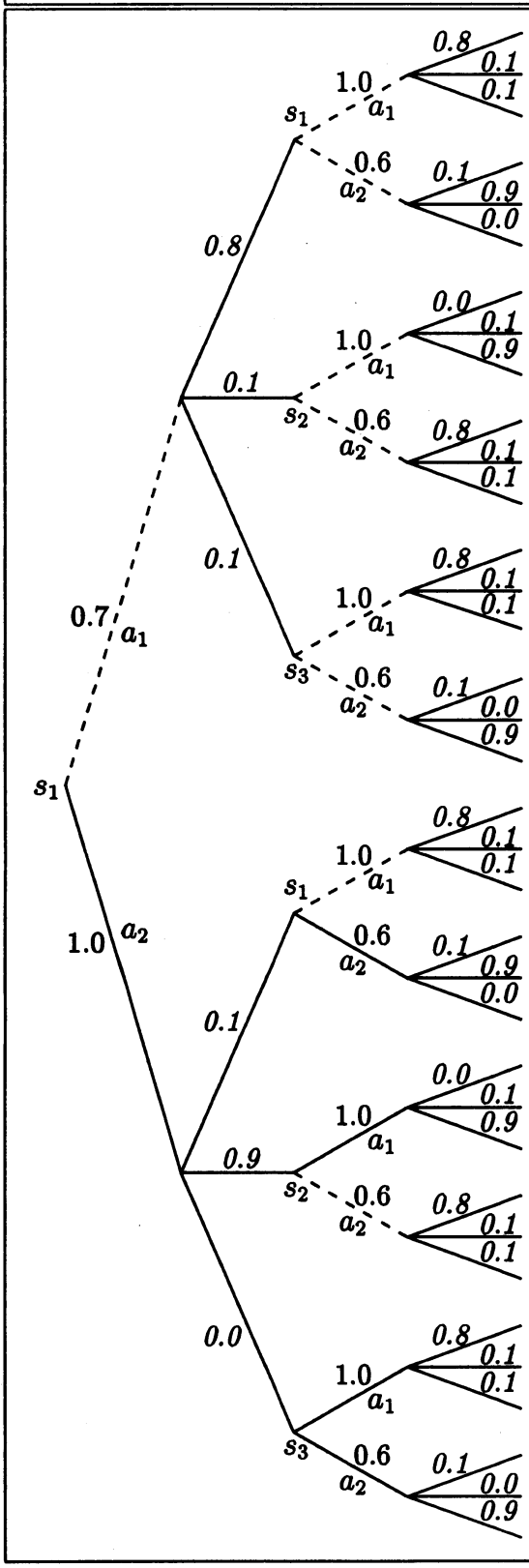
$$\begin{aligned}
&\tilde{\sigma}_0(s_2) = a_2; \qquad \tilde{\sigma}_1(s_2, s_1) = a_2, \quad \tilde{\sigma}_1(s_2, s_2) = a_1, \quad \tilde{\sigma}_1(s_2, s_3) = a_1\\
&\tilde{\sigma}_0(s_3) = a_1; \qquad \tilde{\sigma}_1(s_3, s_1) = a_1, \quad \tilde{\sigma}_1(s_3, s_2) = a_1, \quad \tilde{\sigma}_1(s_3, s_3) = a_1\\
&\qquad\qquad w_0(s_2) = \mathbf{0.84}, \qquad w_0(s_3) = \mathbf{0.28}
\end{aligned}$$

になる。この一般最適政策 $\tilde{\sigma} = \{\tilde{\sigma}_1, \tilde{\sigma}_2\}$ はマルコフでない：$\tilde{\sigma}_1(s_1, s_1) \neq \tilde{\sigma}_1(s_3, s_1)$.

三つの最適な一般政策 σ^*, $\hat{\sigma}$, $\tilde{\sigma}$ は本質的に一致している（$\tilde{\sigma}_1(s_1, s_3) = a_1, a_2$ のいずれでも 2.5 以上になる確率は 0 である）。

$$w_0(s_1) = \max_{\mu} P^{\mu}_{s_1}(\, r_0(U_0) + r_1(U_1) + r_2(X_2) \geq 2.5\,)$$

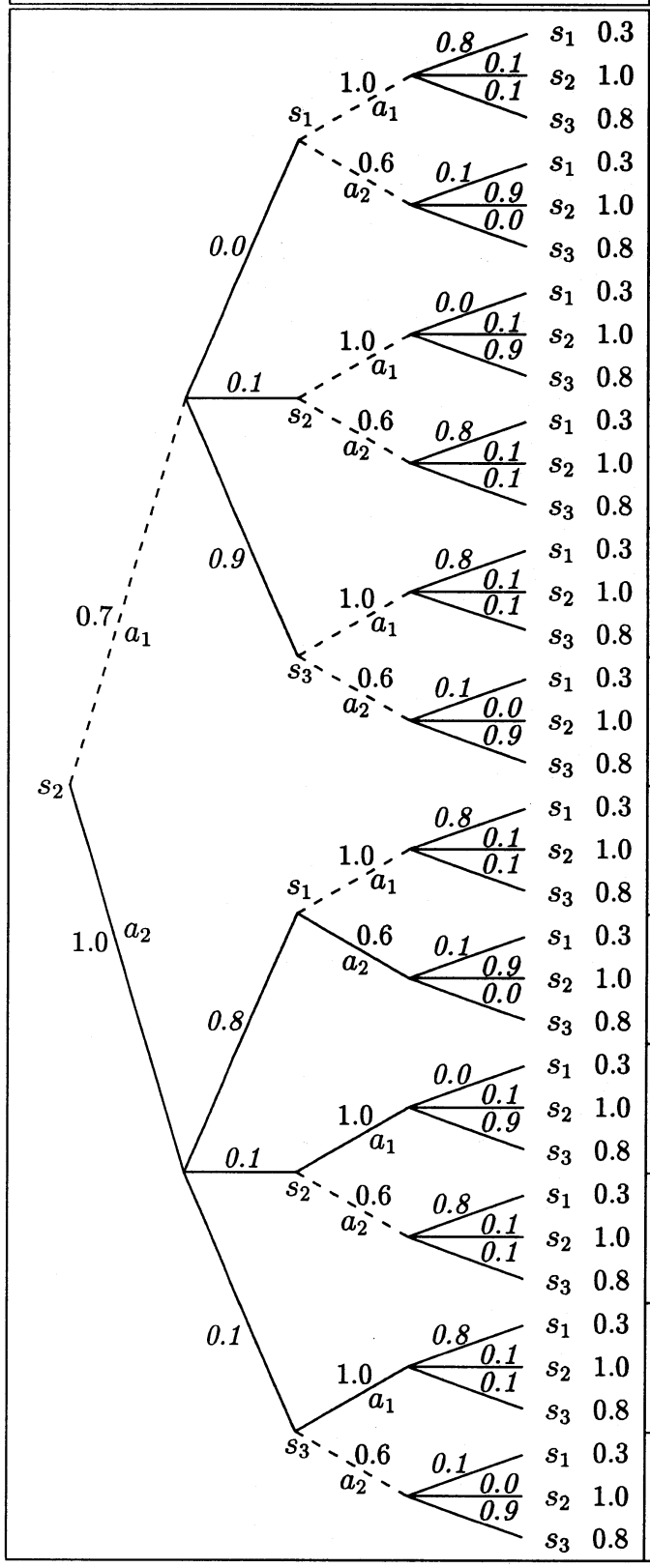
図1：状態 s_1 からの2段確率決定樹表



履歴								加法型評価	経路確率	閾値確率	部分確率	全確率
x_0	r_0/u_0	p_0	x_1	r_1/u_1	p_1	x_2	r_2					
s_1	a_1	0.7	s_1	1.0	0.8	s_1	0.3	2.0	*0.64*	0		
					0.1	s_2	1.0	2.7	*0.08*	0.08	**0.16**	
					0.1	s_3	0.8	2.5	*0.08*	0.08		
				0.6 a_2	0.1	s_1	0.3	1.6	*0.08*	0		
					0.9	s_2	1.0	2.3	*0.72*	0	0	
					0.0	s_3	0.8	2.1	*0.0*	0		
			s_2	1.0 a_1	0.0	s_1	0.3	2.0	*0.0*	0		
					0.1	s_2	1.0	2.7	*0.01*	0.01	**0.1**	
					0.9	s_3	0.8	2.5	*0.09*	0.09		0.28
				0.6 a_2	0.8	s_1	0.3	1.6	*0.08*	0		
					0.1	s_2	1.0	2.3	*0.01*	0	0	
					0.1	s_3	0.8	2.1	*0.01*	0		
			s_3	1.0 a_1	0.8	s_1	0.3	2.0	*0.08*	0		
					0.1	s_2	1.0	2.7	*0.01*	0.01	**0.02**	
					0.1	s_3	0.8	2.5	*0.01*	0.01		
				0.6 a_2	0.1	s_1	0.3	1.6	*0.01*	0		
					0.0	s_2	1.0	2.3	*0.0*	0	0	
					0.9	s_3	0.8	2.1	*0.09*	0		
	a_2	1.0	s_1	1.0 a_1	0.8	s_1	0.3	2.3	*0.08*	0		
					0.1	s_2	1.0	3.0	*0.01*	0.01	0.02	
					0.1	s_3	0.8	2.8	*0.01*	0.01		
				0.6 a_2	0.1	s_1	0.3	1.9	*0.01*	0		
					0.9	s_2	1.0	2.6	*0.09*	0.09	**0.09**	
					0.0	s_3	0.8	2.4	*0.0*	0		
			s_2	1.0 a_1	0.0	s_1	0.3	2.3	*0.0*	0		
					0.1	s_2	1.0	3.0	*0.09*	0.09	**0.9**	
					0.9	s_3	0.8	2.8	*0.81*	0.81		**0.99**
				0.6 a_2	0.8	s_1	0.3	1.9	*0.72*	0		
					0.1	s_2	1.0	2.6	*0.09*	0.09	0.09	
					0.1	s_3	0.8	2.4	*0.09*	0		
			s_3	1.0 a_1	0.8	s_1	0.3	2.3	*0.0*	0		
					0.1	s_2	1.0	3.0	*0.0*	0.0	**0**	
					0.1	s_3	0.8	2.8	*0.0*	0.0		
				0.6 a_2	0.1	s_1	0.3	1.9	*0.0*	0		
					0.0	s_2	1.0	2.6	*0.0*	0.0	**0**	
					0.9	s_3	0.8	2.4	*0.0*	0		

$$w_0(s_2) = \operatorname*{Max}_{\mu} P^{\mu}_{s_1}(\, r_0(U_0) + r_1(U_1) + r_2(X_2) \geq 2.5\,)$$

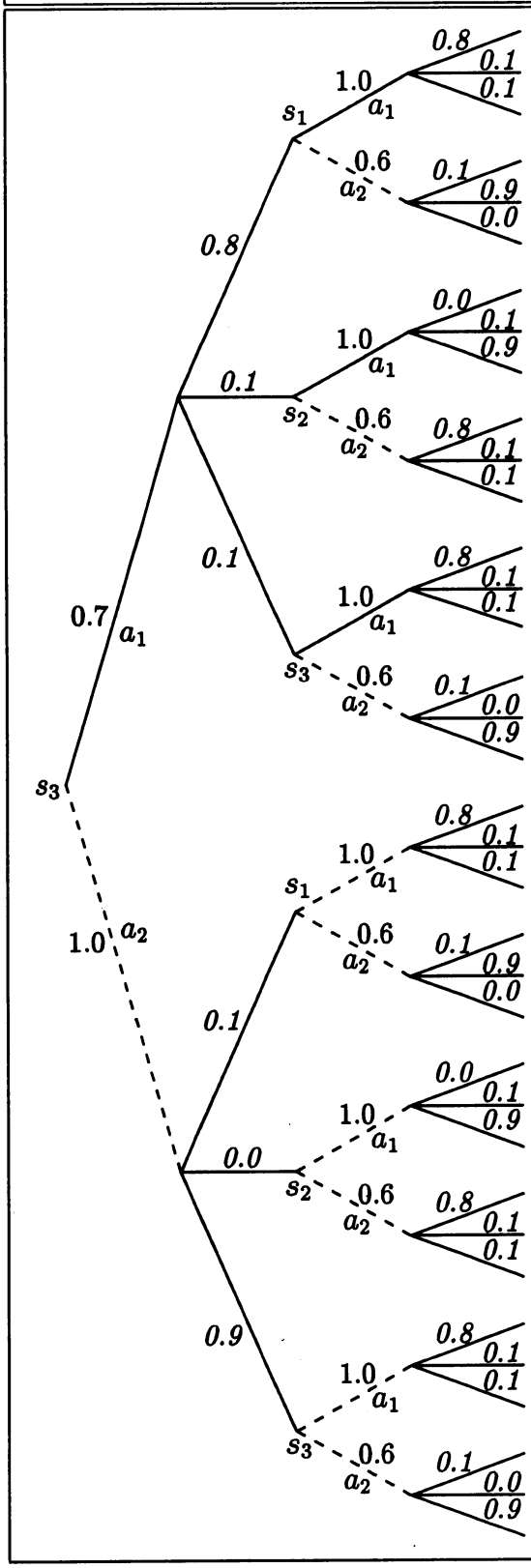
図 2：状態 s_2 からの 2 段確率決定樹表



履歴								加法型評価	経路確率	閾値確率	部分確率	全確率
x_0	r_0/u_0	p_0	x_1	r_1/u_1	p_1	x_2	r_2					
s_2	a_1	0.7	s_1	a_1	1.0	s_1	0.3	2.0	*0.0*	0		0.28
						s_2	1.0	2.7	*0.0*	0.0	**0**	
						s_3	0.8	2.5	*0.0*	0.0		
				a_2	0.6	s_1	0.3	1.6	*0.0*	0		
						s_2	1.0	2.3	*0.0*	0	**0**	
						s_3	0.8	2.1	*0.0*	0		
			s_2	a_1	1.0	s_1	0.3	2.0	*0.0*	0		
						s_2	1.0	2.7	*0.01*	0.01	**0.1**	
						s_3	0.8	2.5	*0.09*	0.09		
				a_2	0.6	s_1	0.3	1.6	*0.08*	0		
						s_2	1.0	2.3	*0.01*	0	0	
						s_3	0.8	2.1	*0.01*	0		
			s_3	a_1	1.0	s_1	0.3	2.0	*0.72*	0		
						s_2	1.0	2.7	*0.09*	0.09	**0.18**	
						s_3	0.8	2.5	*0.09*	0.09		
				a_2	0.6	s_1	0.3	1.6	*0.09*	0		
						s_2	1.0	2.3	*0.0*	0	0	
						s_3	0.8	2.1	*0.81*	0		
	a_2	1.0	s_1	a_1	1.0	s_1	0.3	2.3	*0.64*	0		**0.84**
						s_2	1.0	3.0	*0.08*	0.08	0.16	
						s_3	0.8	2.8	*0.08*	0.08		
				a_2	0.6	s_1	0.3	1.9	*0.08*	0		
						s_2	1.0	2.6	*0.72*	0.72	**0.72**	
						s_3	0.8	2.4	*0.0*	0		
			s_2	a_1	1.0	s_1	0.3	2.3	*0.0*	0		
						s_2	1.0	3.0	*0.01*	0.01	**0.1**	
						s_3	0.8	2.8	*0.09*	0.09		
				a_2	0.6	s_1	0.3	1.9	*0.08*	0		
						s_2	1.0	2.6	*0.01*	0.01	0.01	
						s_3	0.8	2.4	*0.01*	0		
			s_3	a_1	1.0	s_1	0.3	2.3	*0.08*	0		
						s_2	1.0	3.0	*0.01*	0.01	**0.02**	
						s_3	0.8	2.8	*0.01*	0.01		
				a_2	0.6	s_1	0.3	1.9	*0.01*	0		
						s_2	1.0	2.6	*0.0*	0.0	0	
						s_3	0.8	2.4	*0.09*	0		

$$w_0(s_3) = \underset{\mu}{\mathrm{Max}}\, P^{\mu}_{s_1}\left(r_0(U_0) + r_1(U_1) + r_2(X_2) \geq 2.5 \right)$$

図 3：状態 s_3 からの 2 段確率決定樹表



履歴 x_0	r_0/u_0	p_0	x_1	r_1/u_1	p_1	x_2	r_2	加法型評価	経路確率	閾値確率	部分確率	全確率
s_3	a_1	0.7	s_1	a_1	1.0	s_1	0.3	2.0	*0.64*	0		
						s_2	1.0	2.7	*0.08*	0.08	**0.16**	
						s_3	0.8	2.5	*0.08*	0.08		
				a_2	0.6	s_1	0.3	1.6	*0.08*	0		
						s_2	1.0	2.3	*0.72*	0	0	
						s_3	0.8	2.1	*0.0*	0		
			s_2	a_1	1.0	s_1	0.3	2.0	*0.0*	0		
						s_2	1.0	2.7	*0.01*	0.01	**0.1**	
						s_3	0.8	2.5	*0.09*	0.09		**0.28**
				a_2	0.6	s_1	0.3	1.6	*0.08*	0		
						s_2	1.0	2.3	*0.01*	0	0	
						s_3	0.8	2.1	*0.01*	0		
			s_3	a_1	1.0	s_1	0.3	2.0	*0.08*	0		
						s_2	1.0	2.7	*0.01*	0.01	**0.02**	
						s_3	0.8	2.5	*0.01*	0.01		
				a_2	0.6	s_1	0.3	1.6	*0.01*	0		
						s_2	1.0	2.3	*0.0*	0	0	
						s_3	0.8	2.1	*0.09*	0		
	a_2	1.0	s_1	a_1	1.0	s_1	0.3	2.3	*0.08*	0		
						s_2	1.0	3.0	*0.01*	0.01	0.02	
						s_3	0.8	2.8	*0.01*	0.01		
				a_2	0.6	s_1	0.3	1.9	*0.01*	0		
						s_2	1.0	2.6	*0.09*	0.09	**0.09**	
						s_3	0.8	2.4	*0.0*	0		
			s_2	a_1	1.0	s_1	0.3	2.3	*0.0*	0		
						s_2	1.0	3.0	*0.0*	0.0	**0**	
						s_3	0.8	2.8	*0.0*	0.0		0.27
				a_2	0.6	s_1	0.3	1.9	*0.0*	0		
						s_2	1.0	2.6	*0.0*	0.0	**0**	
						s_3	0.8	2.4	*0.0*	0		
			s_3	a_1	1.0	s_1	0.3	2.3	*0.72*	0		
						s_2	1.0	3.0	*0.09*	0.09	**0.18**	
						s_3	0.8	2.8	*0.09*	0.09		
				a_2	0.6	s_1	0.3	1.9	*0.09*	0		
						s_2	1.0	2.6	*0.0*	0.0	0	
						s_3	0.8	2.4	*0.81*	0		

$$
\begin{aligned}
J^0(x_0;\pi) &= P^{\pi}_{x_0}(r_0+r_1+r_2 \geq 2.5) \\
&= E^{\pi}_{x_0}[\psi(r_0+r_1+r_2)] \\
&= \sum_{(x_1,x_2)\in X\times X} \{[\psi(\,r_0(u_0)+r_1(u_1)+r_2(x_2)\,)]p(x_1|\,x_0,u_0)p(x_2|\,x_1,u_1)\}
\end{aligned}
\qquad \text{ただし、} \psi(y)=\begin{cases} 1 & y\geq 2.5 \\ 0 & y<2.5 \end{cases}
$$

表９：全閾値確率ベクトル $J^0(\pi)$, ただし、$\pi=\{\pi_0, \pi_1\}$ はマルコフ政策

π_0 \ π_1	$\begin{pmatrix}a_1\\a_1\\a_1\end{pmatrix}$	$\begin{pmatrix}a_1\\a_1\\a_2\end{pmatrix}$	$\begin{pmatrix}a_1\\a_2\\a_1\end{pmatrix}$	$\begin{pmatrix}a_1\\a_2\\a_2\end{pmatrix}$	$\begin{pmatrix}a_2\\a_1\\a_1\end{pmatrix}$	$\begin{pmatrix}a_2\\a_1\\a_2\end{pmatrix}$	$\begin{pmatrix}a_2\\a_2\\a_1\end{pmatrix}$	$\begin{pmatrix}a_2\\a_2\\a_2\end{pmatrix}$
$\begin{pmatrix}a_1\\a_1\\a_1\end{pmatrix}$	$\begin{pmatrix}0.28\\0.28\\0.28\end{pmatrix}$	$\begin{pmatrix}0.26\\0.10\\0.26\end{pmatrix}$	$\begin{pmatrix}0.18\\0.18\\0.18\end{pmatrix}$	$\begin{pmatrix}0.16\\0.0\\0.16\end{pmatrix}$	$\begin{pmatrix}0.12\\0.28\\0.12\end{pmatrix}$	$\begin{pmatrix}0.10\\0.10\\0.10\end{pmatrix}$	$\begin{pmatrix}0.02\\0.18\\0.02\end{pmatrix}$	$\begin{pmatrix}0.0\\0.0\\0.0\end{pmatrix}$
$\begin{pmatrix}a_1\\a_1\\a_2\end{pmatrix}$	$\begin{pmatrix}0.28\\0.28\\0.20\end{pmatrix}$	$\begin{pmatrix}0.26\\0.10\\0.02\end{pmatrix}$	$\begin{pmatrix}0.18\\0.18\\0.20\end{pmatrix}$	$\begin{pmatrix}0.16\\0.0\\0.02\end{pmatrix}$	$\begin{pmatrix}0.12\\0.28\\0.27\end{pmatrix}$	$\begin{pmatrix}0.10\\0.10\\0.09\end{pmatrix}$	$\begin{pmatrix}0.02\\0.18\\0.27\end{pmatrix}$	$\begin{pmatrix}0.0\\0.0\\0.09\end{pmatrix}$
$\begin{pmatrix}a_1\\a_2\\a_1\end{pmatrix}$	$\begin{pmatrix}0.28\\0.28\\0.28\end{pmatrix}$	$\begin{pmatrix}0.26\\0.26\\0.26\end{pmatrix}$	$\begin{pmatrix}0.18\\0.19\\0.18\end{pmatrix}$	$\begin{pmatrix}0.16\\0.17\\0.16\end{pmatrix}$	$\begin{pmatrix}0.12\\0.84\\0.12\end{pmatrix}$	$\begin{pmatrix}0.10\\0.82\\0.10\end{pmatrix}$	$\begin{pmatrix}0.02\\0.75\\0.02\end{pmatrix}$	$\begin{pmatrix}0.0\\0.73\\0.0\end{pmatrix}$
$\begin{pmatrix}a_1\\a_2\\a_2\end{pmatrix}$	$\begin{pmatrix}0.28\\0.28\\0.20\end{pmatrix}$	$\begin{pmatrix}0.26\\0.26\\0.02\end{pmatrix}$	$\begin{pmatrix}0.18\\0.19\\0.20\end{pmatrix}$	$\begin{pmatrix}0.16\\0.17\\0.02\end{pmatrix}$	$\begin{pmatrix}0.12\\0.84\\0.27\end{pmatrix}$	$\begin{pmatrix}0.10\\0.82\\0.09\end{pmatrix}$	$\begin{pmatrix}0.02\\0.75\\0.27\end{pmatrix}$	$\begin{pmatrix}0.0\\0.73\\0.09\end{pmatrix}$
$\begin{pmatrix}a_2\\a_1\\a_1\end{pmatrix}$	$\begin{pmatrix}0.92\\0.28\\0.28\end{pmatrix}$	$\begin{pmatrix}0.92\\0.10\\0.26\end{pmatrix}$	$\begin{pmatrix}0.11\\0.18\\0.18\end{pmatrix}$	$\begin{pmatrix}0.11\\0.0\\0.16\end{pmatrix}$	$\begin{pmatrix}0.99\\0.28\\0.12\end{pmatrix}$	$\begin{pmatrix}0.99\\0.10\\0.10\end{pmatrix}$	$\begin{pmatrix}0.18\\0.18\\0.02\end{pmatrix}$	$\begin{pmatrix}0.18\\0.0\\0.0\end{pmatrix}$
$\begin{pmatrix}a_2\\a_1\\a_2\end{pmatrix}$	$\begin{pmatrix}0.92\\0.28\\0.20\end{pmatrix}$	$\begin{pmatrix}0.92\\0.10\\0.02\end{pmatrix}$	$\begin{pmatrix}0.11\\0.18\\0.20\end{pmatrix}$	$\begin{pmatrix}0.11\\0.0\\0.02\end{pmatrix}$	$\begin{pmatrix}0.99\\0.28\\0.27\end{pmatrix}$	$\begin{pmatrix}0.99\\0.10\\0.09\end{pmatrix}$	$\begin{pmatrix}0.18\\0.18\\0.27\end{pmatrix}$	$\begin{pmatrix}0.18\\0.0\\0.09\end{pmatrix}$
$\begin{pmatrix}a_2\\a_2\\a_1\end{pmatrix}$	$\begin{pmatrix}0.92\\0.28\\0.28\end{pmatrix}$	$\begin{pmatrix}0.92\\0.26\\0.26\end{pmatrix}$	$\begin{pmatrix}0.11\\0.19\\0.18\end{pmatrix}$	$\begin{pmatrix}0.11\\0.17\\0.16\end{pmatrix}$	$\begin{pmatrix}0.99\\0.84\\0.12\end{pmatrix}$	$\begin{pmatrix}0.99\\0.82\\0.10\end{pmatrix}$	$\begin{pmatrix}0.18\\0.75\\0.02\end{pmatrix}$	$\begin{pmatrix}0.18\\0.73\\0.0\end{pmatrix}$
$\begin{pmatrix}a_2\\a_2\\a_2\end{pmatrix}$	$\begin{pmatrix}0.92\\0.28\\0.20\end{pmatrix}$	$\begin{pmatrix}0.92\\0.26\\0.02\end{pmatrix}$	$\begin{pmatrix}0.11\\0.19\\0.20\end{pmatrix}$	$\begin{pmatrix}0.11\\0.17\\0.02\end{pmatrix}$	$\begin{pmatrix}0.99\\0.84\\0.27\end{pmatrix}$	$\begin{pmatrix}0.99\\0.82\\0.09\end{pmatrix}$	$\begin{pmatrix}0.18\\0.75\\0.27\end{pmatrix}$	$\begin{pmatrix}0.18\\0.73\\0.09\end{pmatrix}$

一般に、加法型評価系であっても閾値確率最適化問題ではマルコフ政策が最適になるとは限らない。事実、表９における（全64個の）マルコフ政策クラスには最適政策は存在しない [2,14]。

参考文献

[1] R.E. Bellman and L.A. Zadeh, Decision-making in a fuzzy environment, *Management Science*, **17**, 1970

[2] M. Bouakiz and Y. Kebir, Target-level criterion in Markov decision processes, *Journal of Optimization Theory and Applications*, **86** (1995), 1-15.

[3] G. H. Hardy, J. E. littlewood and G. Pólya, *Inequalities, 2nd ed.*, Cambridge Univ. Press, 1952.

[4] 岩本 誠一, 「動的計画の最近の進歩」, 第２回ＲＡＭＰシンポジウム論文集, 129–140, 1990.

[5] 岩本 誠一, 多段確率決定樹表について, 日本 OR 学会秋季研究発表会アブストラクト集, pp. 58-59, 1999.

[6] 岩本誠一, 確率最適化における再帰式と決定樹表, 「不確実、不確定性の下での数理的決定理論」京大数理研講究録 **1132**, 2000 年, 15-23.

[7] S. Iwamoto, Maximizing threshold probability through invariant imbedding, *Eds. H.F. Wang and U.P. Wen, Proceedings of The Eighth BELLMAN CONTINUUM*, Hsinchu, ROC, Dec.2000, pp.17-22.

[8] S. Iwamoto, Fuzzy decision-making through three dynamic programming approaches, *Eds. H.F. Wang and U.P. Wen, Proceedings of The Eighth BELLMAN CONTINUUM*, Hsinchu, ROC, Dec.2000, pp.23–27.

[9] S. Iwamoto and T. Fujita, Stochastic decision-making in a fuzzy environment, *J. Operations Res. Soc. Japan*, **38**, 1995

[10] S. Iwamoto, T. Ueno and T. Fujita, Controlled Markov chains with utility functions, *Proc. of The International Workshop on Markov Processes and Controlled Markov Chains*, Changsha, China, August, 1999, in press.

[11] 植野 貴之, マルコフ連鎖上の閾値確率制御, 九州大学大学院経済学研究科修士論文, 2001 年 3 月.

[12] 植野 貴之・岩本 誠一, 最小型評価系の閾値確率制御, 日本 OR 学会秋季研究発表会アブストラクト集, pp. 124-125, 2000.

[13] 植野 貴之・岩本 誠一, 制御マルコフ連鎖上での閾値確率最適化の方法, 研究集会「不確実性の下での数理モデルの構築と最適化」, 2000 年 11 月, 京都大・数理研.

[14] C. Wu and Y. Lin, Minimizing risk models in Markov decision processed with policies depending on target values, *Journal of Mathematical Analysis and Applications*, **231** (1999),